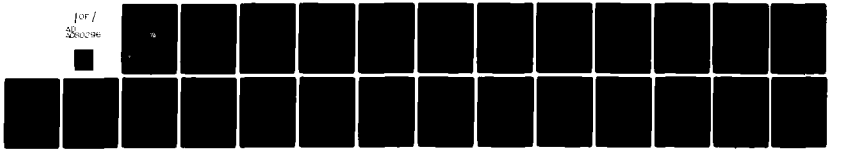
AD-A080 296 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
ON THE CHOICE OF VARIANCE FOR THE LOG RANK TEST.(U)
NOV 79 M BROWN F49620-79-C-0157
UNCLASSIFIED FSU-STATISTICS-M528 AFOSR-TR-80-0068 NL



END
DATE FILMED
-80
DDC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR-80-0068	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): ON THE CHOICE OF VARIANCE FOR THE LOG RANK TEST	5. TYPE OF REPORT & PERIOD COVERED: Interim reptg
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Mark Brown	8. CONTRACT OR GRANT NUMBER(s): F49620-79-C-0157
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Florida State University, Department of Statistics, Tallahassee, FL 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, D. C. 20332	12. REPORT DATE: November 1979
	13. NUMBER OF PAGES: 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Censored data, log rank test, comparison of survival curves, variance estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The log rank test is widely used for comparison of survival curves. This paper examines various estimators for the variance of the log rank test statistic. These include the Mantel-Haenszel variance, the permutation variance of Peto and Peto and several newly proposed estimators. The results generalize to a wide class of test statistics for the two sample problem with censored data.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

ON THE CHOICE OF VARIANCE FOR THE LOG RANK TEST.

by

Mark Brown

FSU Statistics Report No, M528
AFOSR Technical Report No. 79-B3

November, 1979

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research supported by Air Force Office of Scientific Research
Grant No. F49620-79-C-0157.

ON THE CHOICE OF VARIANCE FOR THE LOG RANK TEST.

by

Mark Brown

ABSTRACT

The log rank test is widely used for comparison of survival curves. This paper examines various estimators for the variance of the log rank test statistic. These include the Mantel-Haenszel variance, the permutation variance of Peto and Peto and several newly proposed estimators. The results generalize to a wide class of test statistics for the two sample problem with censored data.

1. Introduction. The log rank test is widely used for comparison of survival curves. There are two versions of the test which differ in the choice of variance for the same test statistic T. The MH(Mantel-Haenszel) variance, an estimate of Var T which is unbiased independent of differences in censoring, and the permutation variance of Peto and Peto, for which the variance is computed under the assumption of equal censoring.

This paper examines the choice of variance. It is observed that for equal sample sizes the MH variance tends to underestimate the true variance when T is large in absolute value. As this is the situation where rejection of the null hypothesis is to be seriously considered, it appears that the true type 1 error will tend to be higher than that assumed, and the reported P value will tend to be too low. Examples with small sample sizes are given to illustrate this phenomena. Further work is needed to see if P values are seriously distorted for larger samples.

In section 4 it is shown that under equal sample sizes the permutation variance tends to overestimate the true variance when censoring is unequal. It thus appears prudent to use the permutation variance rather than the MH variance in assessing significance.

In section 5 alternative variance estimators are proposed.

The log rank results generalize to a wider class of test statistics. This point is discussed in section 6.

2. Review of the log rank test. Sample 1 consists of n individuals and sample 2 of m individuals. For each we observe either the time of death or censoring whichever occurs first, and we know which of the two takes place. For simplicity we assume that no ties occur among the K uncensored observations in the combined sample which are labeled as $t_1 < t_2 \ldots < t_K$. Define R_i to be the number of patients in the combined sample at risk at time t_i, i.e. those for whom the minimum of the death and censoring times are $\geq t_i$. Similarly $R_{1i}(R_{2i})$ is the number of patients at risk from group 1(2) at t_i. Define $\delta_i = 1$ if the death occuring at t_i is from sample 1, 0 if from sample 2. Define $N_1 = \sum_1^K \delta_i$, the number of uncensored observations from sample 1. The hypothesis of interest is H_0: The survival distributions of groups 1 and 2 are the same. The alternative can be one or two-sided, the one-sided alternative favoring group 2 survival would be H_1: The hazard rate for group 1, h_1, exceeds that for group 2, h_2, in the range of interest.

The log rank test statistic is defined by:

$$T = \sum_{i=1}^{K} \left(\delta_i - \frac{R_{1i}}{R_i} \right) = N_1 - \sum_{i=1}^{K} \frac{R_{1i}}{R_i} \tag{2.1}$$

If censoring is assumed to be independent of survival then independent of differences in censoring for the two groups we have:

$$E_{H_0} T = 0 \tag{2.2}$$

$$E_{H_1} T > 0 \tag{2.3}$$

The test of H_0 vs H_1 under the MH(Mantel-Haenszel) approach is to reject H_0 for T large: the distribution of T under H_0 is approximated by a normal distribution with mean 0 and variance:

(2.4) $$\sigma^2_{MH} = \sum_{i=1}^{K} \frac{R_{1i}R_{2i}}{R_i^2}$$

The MH variance, σ^2_{MH}, is an unbiased estimator of $\mathrm{Var}_{H_0} T$, independent of differences in censoring between the two groups.

Peto and Peto [10] observed that under the assumption of equal censoring a test based on the log rank statistic, T, can be interpreted as a sum of scores rank test. Define the score of a patient with observation Z_ℓ as $U_\ell = \epsilon_\ell - \sum_{i:t_i \le Z_\ell} \frac{1}{R_i}$, where $\epsilon_\ell = \begin{cases} 1 \text{ if } Z_\ell \text{ is uncensored} \\ 0 \text{ if } Z_\ell \text{ is censored.} \end{cases}$ Then $T = \sum_{\text{group 1}} U_\ell$.

The permutation variance of T is easily computed as:

(2.4) $$\sigma^2_p = \frac{nm}{(n+m)(n+m-1)} \left[K - \sum_{i=1}^{K} \frac{1}{R_i} \right]$$

The permutation test based on T is a conditional test given A, where A is the sample outcome with group identity of patients deleted. In the absence of equal censoring neither the MH test nor the permutation test based on T are conditional tests given A. The conditional mean of T under H_0 given A depends on the censoring and will generally not be zero. Furthermore σ^2_{MH} is generally not conditionally unbiased for $\mathrm{Var}_{H_0}(T|A)$.

3. A discussion of the M-H variance. Suppose that $n = m$. Below is a plausibility argument for the assertion that the MH variance tends to underestimate the true variance, $\mathrm{Var}_{H_0} T$, when T is large in absolute value. A consequence of this would be exageratedly small P values under the MH test. The plausibility argument is difficult to quantify. Illustrative examples with small sizes are given, which do indicate the existence of a problem. Further work is needed to examine how serious the problem is for larger sample sizes.

Consider the sample points which lead to a large absolute value of T, and thus lead the statistician to consider rejection of H_0. In general they are characterized by an excess of uncensored observations from one of the samples. This leads for many i to an inbalance between R_{1i} and R_{2i} causing $\frac{R_{1i} R_{2i}}{R_i^2}$ to be small, and consequently σ_{MH}^2 to be small. Large values of $|T|$ tend to be accompanied by small values of σ_{MH} causing exageratedly large $|T|/\sigma_{MH}$, and P values which are too small. The statistician is observing $\frac{R_{1i}}{R_i}$ close to zero or one, seeing that it has resulted largely from an inbalance in deaths rather than censoring, and then estimating the variance at that stage by $\frac{R_{1i} R_{2i}}{R_i^2}$, is ignoring the fact that under H_0 $\frac{R_{1i}}{R_i}$ should not be as extreme as in the current case. As the statistician is attempting to estimate the variance under H_0, he is ignoring an important consideration. The fact that in other cases ($|T|$ small) ignoring similar information leads to an overestimate, and the two balance out to make σ_{MH}^2 unbiased, does not alleviate the difficulty.

Below are two cases, $n = m = 3$, and $n = m = 4$, both without censoring.

Table 1 below considers $n = m = 3$. The values of σ^2_{MH} appear in column 1. In column 2 the average value of $|T|$ is computed for sample points with common σ^2_{MH}. Note that as σ^2_{MH} increases $E[|T||\sigma^2_{MH}]$ decreases. The sample correlation coefficient between σ^2_{MH} and $|T|$ equals -.86. The average of σ^2_{MH} over the 20 sample points equals the permutation variance 1.065. If the ranks 1, 2, 3 are observed for sample 1 then the exact P value is .05; the P value using a normal approximation with permutation variance is .0365; the P value using $\sigma^2_{MH} = .68$ as approximating variance is .0124.

Table 1 (n=m=3; no censoring)

σ^2_{MH}	$E(\|T\|\|\sigma^2_{MH})$	number of cases
.68	1.85	2
.90	1.52	2
.96	1.17	4
1.15	.52	4
1.21	.50	8

Table 2 is similar to table 1 with $n = m = 4$. In this case the sample correlation coefficient between σ^2_{MH} and $|T|$ equals -.88, slightly higher in absolute value than for $n = m = 3$. The permutation variance equals 1.51.

Table 2 (n=m=4; no censoring)

σ^2_{MH}	$E(\|T\| \mid \sigma^2_{MH})$	number of cases
.88	2.54	2
1.06	2,29	2
1.14	2.09	2
1.17	1.85	4
1.29	1.95	2
1.37	1.75	2
1.39	1.52	4
1.43	1.17	4
1.46	1.17	8
1.54	.95	4
1.62	.75	4
1.64	.54	8
1.68	.50	8
1.71	.50	16

Table 3, again for n = m =4 without censoring, lists the 7 largest values of T with corresponding exact P values $\frac{1}{70}$, $\frac{2}{70}$, ..., $\frac{7}{70}$, and P values under the normal approximations with both MH and permutation variance. It is seen that the P values are systematically too low using the MH variance.

Table 3

value of T	Exact P value	σ^2_{MH} P value	σ^2_P P value
2.54	.0143	.0035	.0194
2.29	.0286	.0133	.0312
2.09	.0429	.0255	.0446
1.95	.0571	.0424	.0558
1.92	.0714	.0380	.0592
1.78	.0857	.0502	.0738
1.75	.1000	.0667	.0770

The fact that in table 3 the permutation variance gives a better approximation is not surprising. We are computing the exact P value under the conditional distribution of T given A, under the assumption of equal censoring. This is the "home court" for the permutation variance. At this point we are not arguing the virtue of the permutation variance, but are illustrating the behavior of σ^2_{MH} described earlier.

Note that when n = m and censoring is equal a simple symmetry argument shows that $Cov(\sigma^2_{MH}, T) = 0$. This is an example for which two random variables are uncorrelated but highly dependent.

If one could show that the MH variance is the conditional variance of T given an appropriate function of the data for which the conditional mean of T remains zero, then the above mentioned behavior of σ^2_{MH} would no longer be disturbing. The MH variance would be the appropriate variance for a well defined conditional experiment. No such interpretation has been offered however and it is clear by simple counterexamples that any such interpretation would have to be asymptotic rather than exact.

4. A property of the permutation variance. Consider the case of equal sample sizes and random censoring, with the group 1 censoring distribution, H, allowed to differ from the group 2 censoring distribution I. Recall that $\sigma_p^2 = \frac{n}{2(2n-1)}\left(K - \sum_1^K \frac{1}{R_i}\right)$, the permutation variance, is a function of $(K,R_1,\ldots,R_K)$ which under equal censoring is the conditional variance of T given $(K,R_1\ldots,R_K)$. In theorem 1 below it is shown that $\sigma_p^2 \geq \tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is a function of $(K,R_1,\ldots,R_K)$ depending on (F,H,I) which is unbiased for $\mathrm{Var}_{H_0} T$. Thus for any (F,H,I), $E_{H_0}\sigma_p^2 \geq E_{H_0}(\tilde{\sigma}^2) = \mathrm{Var}_{H_0} T$.

In view of this property, the fact that σ_p^2 does not vary with T for fixed $(K,R_1,\ldots,R_K)$ as σ_{MH}^2 does, and the difficulty with σ_{MH}^2 discussed in section 3, it would seem prudent under approximately equal sample sizes to use σ_p^2 rather than σ_{MH}^2 in assessing significance.

The following two lemmas are needed in the proof of theorem 1.

Lemma 1. Let $\underline{\alpha} = \begin{pmatrix}\alpha_0\\ \vdots \\ \alpha_m\end{pmatrix}$ and $\underline{\beta} = \begin{pmatrix}\beta_0\\ \vdots \\ \beta_m\end{pmatrix}$ be two probability distributions on $\{0,1,\ldots,m\}$ with $\alpha_0 \leq \beta_0$ and $\frac{\alpha_i}{\alpha_{i-1}} \geq \frac{\beta_i}{\beta_{i-1}}$, $i = 1, \ldots, m$. Then $\underline{\alpha}$ is stochastically larger than $\underline{\beta}$ i.e. $\sum_r^m \alpha_i \geq \sum_r^m \beta_i$, $r = 0, \ldots, m$.

Proof. For any k, $\frac{1}{\alpha_k}\sum_k^m \alpha_j = \sum_{j=k}^m \frac{\alpha_j}{\alpha_k} =$

$1 + \sum_{j=k+1}^m \left(\prod_{k+1}^j \frac{\alpha_i}{\alpha_{i-1}}\right) \geq 1 + \sum_{j=k+1}^m \left(\prod_{k+1}^j \frac{\beta_i}{\beta_{i-1}}\right) = \frac{1}{\beta_k}\sum_k^m \beta_j$. Therefore

$h_k(\beta) = \dfrac{\beta_k}{\sum_k^m \beta_j} \geq \dfrac{\alpha_k}{\sum_k^m \alpha_j} = h_k(\alpha)$, so $\sum_r^m \beta_j = \prod_0^{r-1} (1-h_k(\beta)) \leq \prod_0^{r-1} (1-h_k(\alpha)) = \sum_r^m \alpha_k$. ||

Lemma 2. Let $X \sim B(n,p_1)$, $Y \sim B(n,p_2)$. Then $E[X(r-X)\,|\,X+Y=r]$ is maximized when $p_1 = p_2$. Its maximum value is $\dfrac{nr(r-1)}{2(2n-1)}$.

Proof. (i) Case 1, $r \leq n$. Define:

$$\gamma_i(c) = \begin{cases} \binom{n}{j}\binom{n}{r-j}(c^j + c^{r-j}), & j < \frac{r}{2} \\ \binom{n}{r/2}^2 c^{r/2}, & j = \frac{r}{2},\ r \text{ even} \end{cases} \tag{4.1}$$

and

$$\alpha_i(c) = \frac{\gamma_i(c)}{\sum_{j=0}^{[r/2]} \gamma_j(c)}, \quad i = 0, \ldots, [r/2] \tag{4.2}$$

Note that:

$$E(X(r-X)\,|\,X+Y=r) = \sum_{\ell=0}^{[\frac{r}{2}]} \ell(r-\ell)\alpha_\ell(c) \tag{4.3}$$

where

$$c = \frac{p_1|q_1}{p_2|q_2} \tag{4.4}$$

Since $\ell(r-\ell)$ is increasing in the range $[0,[\frac{r}{2}]]$, (4.1) will be maximized by that value of c(if such a value exists) for which the probability distribution $\alpha(c)$, defined by (4.2), is stochastically largest.

We now prove that $c = 1$ gives the stochastically largest distribution. By lemma 1 it will suffice to prove for abritrary positive c that:

$$\alpha_0(1) \leq \alpha_0(c) \tag{4.5}$$

$$\frac{\alpha_i(1)}{\alpha_{i-1}(1)} \geq \frac{\alpha_i(c)}{\alpha_{i-1}(c)}, \quad i = 1, \ldots, [r/2] \tag{4.6}$$

The statement (4.5) is equivalent to:

$$\frac{1+c^r}{2} \geq \sum_{\ell=0}^{r} \frac{\binom{n}{\ell}\binom{n}{r-\ell}}{\binom{2n}{r}} \left(\frac{c^{\ell}+c^{r-\ell}}{2}\right) \tag{4.7}$$

The right side is a weighted average of $\frac{c^{\ell}+c^{r-\ell}}{2}$ while the left side puts all its weight on the value $\frac{1+c^r}{2}$. Since the function $g(x) = c^x + c^{r-x}$ is decreasing in $[0,r/2)$ and $g(x) = g(r-x)$, it follows that for fixed c g is maximized at $x = 0$ or equivalently at $x = r$. Thus (4.5) is proved.

To prove (4.6) note that:

$$\frac{\alpha_i(c)}{\alpha_{i-1}(c)} = \lambda_i \frac{c^{i}+c^{r-i}}{c^{i-1}+c^{r-i-1}}, \quad \begin{matrix} r \text{ odd, } i = 1, \ldots, \frac{r-1}{2} \\ r \text{ even, } i = 1, \ldots, \frac{r-2}{2} \end{matrix} \tag{4.8}$$

$$\frac{\alpha_i(c)}{\alpha_{i-1}(c)} = \lambda_{r/2} \frac{c^{r/2}}{c^{r/2-1}+c^{r/2+1}}, \quad r \text{ even, } i = r/2 \tag{4.9}$$

where λ_i is a constant independent of c. In both cases ((4.8) and (4.9) a simple differentiation argument shows that the maximum is achieved at $c = 1$. This concludes the proof of case 1.

(ii) Case 2, $r \geq n$. Suppose $r > n$. Then $(n-X) + (n-Y) = 2n - r < n$, thus by case 1, $E[(n-X)[2n-r)-(n-X)]|X+Y=r]$ is maximized by choosing $p_1 = p_2$. But $(n-X)[(2n-r)-(n-X)] = X(r-X)+n(n-r)$. Thus maximizing $E[(n-X)[(2n-r)-(n-X)]|X+Y=r]$ is equivalent to maximizing $E(X(r-X)|X+Y=r)$, and case 2 is proved.

Finally note that when $p_1 = p_2$,

$$\sum_{\ell=0}^{[r/2]} \ell(r-\ell)\alpha_{\ell}(1) = \frac{n^2 \sum_{0}^{r-2} \binom{n-1}{\ell}\binom{n-1}{r-2-\ell}}{\sum_{0}^{r} \binom{n}{\ell}\binom{n}{r-\ell}} = \frac{nr(r-1)}{2(2n-1)} \quad ||.$$

Theorem 1. Assume n = m. Let the censoring variables for sample 1 (2) be i.i.d. with distribution H(I). Assume that the censoring variables are independent of the survival variables and that the two groups are independent of each other. Then:

(4.10) $$E_{H_0}\,\sigma_p^2 \geq \mathrm{Var}_{H_0} T$$

Proof. The proof below is heuristic but can be made rigorous using martingales. For j = 1, 2 define $R_j(u)$ to be the number of group j patients at risk at time u^-; define $R(u) = R_1(u)+R_2(u)$. Define $\mathcal{H}(u)$ to be the history of the process until u^-; this includes all censored and uncensored observations and knowledge of group membership of the patients involved. Define N(t) to be the number of uncensored observations in [0,t], and let h denote the common hazard rate function for the two groups. Note that under H_0, dN(u) and $\mathcal{H}(u)$ are conditionally independent given R(u), with $E(dN(u)\,|\,\mathcal{H}(u)) = E(dN(u)\,|\,R(u)) = R(u)h(u)du$. It follows that:

$$E_{H_0}[g(\mathcal{H}(u))dN(u)] = E_{H_0}[E_{H_0}(g(\mathcal{H}(u))\,|\,R(u))dN(u)]$$

$$= E_{H_0}[E_{H_0}(g(\mathcal{H}(u))\,|\,R(u))R(u)h(u)du]$$

Now $\sigma_{MH}^2 = \int \frac{R_1(u)R_2(u)}{R^2(u)}\,dN(u)$ is unbiased for $\mathrm{Var}_{H_0} T$. Defining $\bar{\sigma}^2 = \int \frac{1}{R^2(u)} E(R_1(u)R_2(u)\,|\,R(u))dN(u)$, and using (4.11) with $g(\mathcal{H}(u)) = R_1(u)R_2(u)$, we obtain:

(4.12) $$E_{H_0}\,\bar{\sigma}^2 = \mathrm{Var}_{H_0} T$$

We wish to compute $\tilde{\sigma}^2$ under known (F,H,I). For convenience we adopt the convention of using left continuous rather than right continuous cdf's. Thus, for example, $H(x) = P_H(X<x)$, and $\bar{H}(x) = P_H(X \geq x)$. Now $R_1(u) \sim B(n, \bar{F}(u)\bar{H}(u))$ and $R_2(u) \sim B(m, \bar{F}(u)\bar{I}(u))$, where B(n,p) denotes the binomial distribution with parameters n and p. By lemma 2, $E_{H_0}(R_1(u)R_2(u)|R(u))$ is maximized when $H(u) = I(u)$, the maximum value being $\frac{n}{2(2n-1)}R(u)(R(u)-1)$.

Thus $\sigma_p^2 = \int \frac{1}{R^2(u)}\left[\frac{n}{2(n-1)}R(u)(R(u)-1)\right]dN(u) \geq \int \frac{1}{R^2(u)}E_{H_0}[R_1(u)R_2(u)|R(u)]dN(u) = \tilde{\sigma}^2$,

and $E_{H_0}\sigma_p^2 \geq E_{H_0}\tilde{\sigma}^2 = \mathrm{Var}_{H_0} T$.

5. Alternative variance estimators. It was shown in section 4 that $\tilde{\sigma}^2$ is unbiased for $\mathrm{Var}_{H_0} T$ and in the case of equal censoring reduces to the permutation variance σ_p^2; under equal censoring σ_p^2 is the appropriate variance for a conditional test given A.

The estimator $\tilde{\sigma}^2$ also works well in the following case. Assume that all group 1 observations are censored at s and all group 2 observations at times which exceed s. Define N(s) to be the number of uncensored observations from the combined groups in [0,s]. Then $\tilde{\sigma}^2$ reduces to:

$$\frac{nm}{(n+m)(n+m-1)} \sum_{1}^{N(s)} \frac{n+m-i}{n+m-i+1} \tag{5.1}$$

Note that (5.1) is unbiased for $\mathrm{Var}_{H_0} T$ and is a function of N(s). Under H_0 the conditional distribution of T given N(s) is independent of F and $E_{H_0}(T|N(s)) = 0$, therefore $\mathrm{Var}_{H_0}(T|N(s))$ is also a function of N(s) which is unbiased for $\mathrm{Var}_{H_0} T$. Since N(s) is complete (follows from completeness of the binomial family) it follows that $\tilde{\sigma}^2 = \mathrm{Var}_{H_0}(T|N(s))$. Thus once again $\tilde{\sigma}^2$ reduces to a conditional variance and is the appropriate variance for a conditional test.

One further example to illustrate the use of $\tilde{\sigma}^2$: Suppose that all patients enter the study at the same time, censoring is due only to termination of the study, and the study terminates at w_k, the time of the k^{th} smallest order statistic from the combined sample. In this case $\tilde{\sigma}^2$ is easily computed to be:

$$\frac{nm}{(n+m)(n+m-1)} \sum_{i=1}^{k} \frac{n+m-i}{n+m-i+1} \tag{5.2}$$

Since (5.2) is unbiased for $\mathrm{Var}_{H_0} T$ and is a constant, it coincides with $\mathrm{Var}_{H_0} T$. This implies that $\mathrm{Var}_{H_0} T$ is independent of F, a fact which can easily be independently verified. Note that in the special cases $k = n+m$ and $k = n+m-1$ (5.2) coincides with σ_p^2.

Note that $\tilde{\sigma}^2$ can only be computed under censoring assumptions. For the three specific cases discussed above, $\tilde{\sigma}^2$ gave useful and appropriate variances.

(5.3) Case of known censoring times. Suppose that all censoring times, including those corresponding to uncensored observations are known. For example, if the data is analyzed at time T and the only cause of censoring is termination of the study, then a patient entering the study at time x has a censoring time of T-x.

In this case we think of the censoring times as fixed numbers. Hypothetical repetitions of the experiment yield different survival times but the same fixed censoring times. The variance of the statistic T under H_0 will depend on F. This variance can be unbiasedly extimated by $\tilde{\sigma}^2$. To do so we need to compute $E_{H_0}(R_1(u)R_2(u)|R(u))$ under known censoring times. Define $\ell_j(u)$, $j = 1, 2$ to be the number of group j patients with censoring time $\geq u$, and $\ell(u) = \ell_1(u)+\ell_2(u)$. There are $\ell(u)$ potential members of the risk set at time u all equally likely under H_0 to be included. The conditional distribution of $R_1(u)$ given R(u) is thus hypergeometric. It easily follows that:

$$E_{H_0}(R_1(u)R_2(u)|R(u) = \frac{\ell_1(u)\ell_2(u)R(u)(R(u)-1)}{\ell(u)(\ell(u)-1)} \tag{5.4}$$

To distinguish the estimator $\tilde{\sigma}^2$ in the case of known censoring from previously considered cases we will denote it by $\hat{\sigma}^2$. Then from (5.4) and (4.11):

(5.5) $$\hat{\sigma}^2 = \sum_{i=1}^{k} \frac{\ell_{1i}\ell_{2i}(R_i-1)}{\ell_i(\ell_i-1)R_i}$$

(5.6) $$E_{H_0}\hat{\sigma}^2 = \mathrm{Var}_{H_0} T$$

where $\ell_{ji} = \ell_j(t_i)$, $j = 1, 2$ and $\ell_i = \ell_{1i} + \ell_{2i}$.

Note that when $\ell_{1i} = R_{1i}$, $\ell_{2i} = R_{2i}$, $i = 1, \ldots, K$, then $\hat{\sigma}^2 = \sigma^2_{MH}$. Thus $\hat{\sigma}^2$ and σ^2_{MH} will be close together when most of the uncensored abservations occur very shortly before their censoring times. Also note that when $\ell_{1i} = n$, $\ell_{2i} = m$ then $\hat{\sigma}^2 = \sigma^2_p$. The reason for this is that under equal censoring all n+m patients have the same probability under H_0 of being in the risk set at time u.

(5.6) Remark. The exact variance in the case of known censoring times is computed by:

$$\mathrm{Var}_{H_0} T = E_{H_0} \int \frac{\ell_1(u)\ell_2(u)}{\ell(u)(\ell(u)-1)} E_{H_0} \left[\frac{R(u)-1}{R(u)} dN(u)\right]$$

$$= \int \frac{\ell_1(u)\ell_2(u)}{\ell(u)(\ell(u)-1)} \left[\ell(u) - \frac{1-(F(u)^{\ell(u)}}{\bar{F}(u)}\right] dF(u).$$

(5.7) Remark. It is undesirable on philosophical grounds to have the statistical procedure depend on potential but unrealized censoring. Pratt [2] has interesting comments on this point and Cox [3] and Efron [5] briefly mention it as well.

The estimator $\hat{\sigma}^2$ depends on unrealized censoring times while σ^2_{MH}, σ^2_p, and σ^{*2} (defined below) do not. Nevertheless, when all censoring times are known I would use $\hat{\sigma}^2$. I would not throw out part of the information just to avoid the above philosophical objection.

It is clear that the values of the unrealized censoring times do not provide intrinsic information about survival, the attribute of interest. However they do provide information about the distribution of the test statistic under hypothetical repititions of the experiment. In a frequentist approach such information is often useful. To a Bayesian or follower of the liklihood principle it is irrelevant.

(5.8) Remark. The rational for $\tilde{\sigma}^2$ is as follows. The problem of estimating $\mathrm{Var}_{H_0} T$ can be reduced in an appropriate sense to that of estimating $\gamma(u) = E_{H_0}[(dN_1(u) - \frac{R_1(u)}{R(u)} dN(u))^2 | dN(u)]$, where R_1, R, N have previously been defined and $N_1(t)$ is the number of group 1 uncensored observations in $[0,t]$. The M-H estimator estimates $\gamma(u)$ by

$$\gamma_{MH}(u) = E_{H_0}\left[\left[dN_1(u) - \frac{R_1(u)}{R(u)}\right]^2 | R_1(u), R_2(u), dN(u)\right];$$ $\tilde{\sigma}^2$ estimates $\gamma(u)$ by

$$\tilde{\gamma}(u) = E_{H_0}(\gamma_{MH}(u) | R(u)).$$

In comparing $\tilde{\gamma}(u)$ to $\gamma_{MH}(u)$ we see that $\mathrm{Var}_{H_0}(\tilde{\gamma}(u)) \le \mathrm{Var}_{H_0}(\gamma_{MH}(u))$ while $\mathrm{Cov}_{H_0}(\tilde{\gamma}(u), \tilde{\gamma}(v)) = \mathrm{Cov}_{H_0}(\gamma_{MH}(u), \gamma_{MH}(u))$; it thus follows that $\mathrm{Var}_{H_0}\tilde{\sigma}^2 \le \mathrm{Var}_{H_0}\sigma^2_{MH}$. To interpret this result properly recall that σ^2_{MH} is computable from the data without censoring assumptions while $\tilde{\sigma}^2$ requires special assumptions. When $\tilde{\sigma}^2$ is computed under specific censoring assumptions then $\mathrm{Var}_{H_0}\tilde{\sigma}^2 \le \mathrm{Var}_{H_0}\sigma^2_{MH}$ under these assumptions, but not necessarily more generally.

While reduction in variance is welcome my main reason for employing $\tilde{\sigma}^2$ is that I believe it will work better than σ^2_{MH} when $|T|$ is large. The difficulty

in estimating $\gamma(u)$ when $|T|$ is large is that we need an estimate of variance under H_0 from data which is extreme under H_0. We do not want to use the observed variance $\gamma_{MH}(u)$ which we know is untypical of H_0. The estimator $\tilde{\gamma}(u)$ removes the dependence of the variance estimator on $R_1(u)/R(u)$ and attempts to replace the observed variance $\gamma_{MH}(u)$ by a quantity which better estimates its expection, $\gamma(u)$.

The computation of $\tilde{\gamma}(u)$ is simple under assumptions which imply that there are a pool of patients who are potential members of the risk set at u, and the R(u) patients who are actually in the risk set are chosen from the pool by random sampling without replacement. This structure characterizes the examples we have so far studied.

(5.9) Random Censoring. If we assume random censoring with distribution H for group 1 and I for group 2, then the variance of T depends on (F,H,I) in a complex way. The unbiased estimator $\tilde{\sigma}^2$ also depends on (F,H,I), specifically:

$$(5.10) \qquad E_{H_0}(R_1(u)R_2(u)\,|\,R(u)=r) = \frac{\sum_{\ell=\alpha}^{\beta} \ell(r-\ell)\binom{n}{\ell}\binom{m}{r-\ell}c_u^{\ell}}{\sum_{\ell=\alpha}^{\beta}\binom{n}{\ell}\binom{m}{r-\ell}c_u^{\ell}}$$

where:

$$(5.11) \qquad \alpha = \min(0,m-r),\ \beta = \max(r,n),\ c(u) = \frac{\bar{H}(u)(1-\bar{F}(u)\bar{I}(u))}{\bar{I}(u)(1-\bar{F}(u)\bar{H}(u))}$$

Rather than try to estimate $\tilde{\sigma}^2$ we opt for a different approach. The estimator $\hat{\sigma}^2$ requires knowledge of ℓ_{1i}, ℓ_{2i}, $i = 1, \ldots, k$. We no longer know the censoring times corresponding to uncensored observations. However, defining $N_{j,i}$ $j = 1, 2$ as the number of group j observations among the first i uncensored observations, we know that $R_{j,i-1} \le \ell_{ji} \le R_{ji} + N_{j,i-1}$, $j = 1, 2$.

This suggests estimating ℓ_{1i} by:

$$\hat{\ell}_{1i} = R_{1i} + \sum_{j<i} \delta_j \frac{\hat{\bar{H}}(t_i)}{\hat{\bar{H}}(t_j)} \tag{5.12}$$

where $\delta_j = 1$ or 0 depending on whether the j^{th} uncensored observation is from group 1 or 2, and $\hat{H}$ is the Kaplan-Meier estimator of the censoring distribution for group 1.

By Efron's ([5]) interesting observation of the self-consistency property of the Kaplan-Meier estimator:

$$\hat{\ell}_{1i} = n\hat{\bar{H}}(t_i) \tag{5.13}$$

$$\hat{\ell}_{2i} = m\hat{\bar{I}}(t_i) \tag{5.14}$$

We now define an estimator σ^{*2} by substituting $\hat{\ell}_{1i}$, $\hat{\ell}_{2i}$, $\hat{\ell}_i = \hat{\ell}_{1i} + \hat{\ell}_{2i}$ for ℓ_{1i}, ℓ_{2i} and ℓ_i in $\hat{\sigma}^2$ (5.5) obtaining:

$$\sigma^{*2} = \sum_{i=1}^{k} \frac{\hat{\ell}_{1i}\hat{\ell}_{2i}(R_i - 1)}{\hat{\ell}_i(\hat{\ell}_i - 1)R_i} \tag{5.15}$$

The estimator σ^{*2} performs well in the following cases. If group 1 observations are censored at s, group 2 observations at times which exceed s then σ^{*2} coincides with (5.1). If censoring occurs at w_k, the k^{th} smallest order statistic in the combined sample, then σ^{*2} coincides with (5.2). In the special case $k = m+n$ (corresponding to no censoring) σ^{*2} equals σ_p^2. If $\hat{H}$ and $\hat{I}$ are approximately equal, indicative of equal censoring, then σ^{*2} will be close to σ_p^2.

The estimator σ^{*2} is an estimate of $\hat{\sigma}^2$ which in turn appears to be superior to σ^2_{MH} (see remark (5.9). It is hoped that σ^{*2} will generally be better than σ^2_{MH}, particularly when $|T|$ is large.

(5.16) Remark. Another expression for $\text{Var}_{H_0} T$ under random censoring is:

$$\text{Var}_{H_0} T = nm \int \bar{H}(u)\bar{I}(u)\bar{F}^2(u) E\left[\frac{R(u)}{B+B'+2}\right] h(u)du \tag{5.17}$$

where $B \sim B(n-1,\bar{H}(u)\bar{F}(u))$, $B' \sim B(m-1,\bar{I}(u)\bar{F}(u))$ with B and B' independent of each other and of R(u). Estimating $n\bar{H}(u)$ by $\hat{\ell}_1(u)$, $m\bar{I}(u)$ by $\hat{\ell}_2(u)$, $\bar{F}^2(u)$ by $\frac{R(u)(R(u)-1)}{\hat{\ell}(u)(\hat{\ell}(u)-1)}$, h(u) by $\frac{dN(u)}{R(u)}$, and approximating $E\left[\frac{R(u)}{B+B'+2}\right]$ by 1 in (5.17) yields the estimator σ^{*2}(5.15).

(5.18) Remark. Both σ^{*2} and σ^2_{MH} throw out observations which are censored prior to t_1, that is they treat the data as if $n = R_{11}$, $m = R_{21}$ and no uncensored observations occur prior to t_i. This is quite reasonable and desirable. In fact attention should be focussed on $\text{Var}_{H_0}(T|R_{11},R_{21})$ rather that $\text{Var}_{H_0} T$, and σ^2_p and $\hat{\sigma}^2$ should be adjusted accordingly.

(5.19) One more censoring situation. Termination of the study occurs at time T_1 for group 1 and T_2 for group 2. Termination is one cause of censoring; other causes are present such as withdrawal from the study, loss to follow-up, etc. A group 1 patient entering the study at time x has a maximum censoring time of $T_1 - x$ but may be censored sooner; similarly for group 2. Define $\hat{Q}_i$, i = 1, 2 to be the Kaplan-Meier estimate of Q_i the group i distribution of censoring due to causes other than termination of the study. In estimating Q_i the observations corresponding to death and censoring due to termination

of the study are treated as censored, those corresponding to other causes of censoring are treated as uncensored. Define:

$$\hat{\bar{Q}}_{1,s}(t) = \begin{cases} \hat{\bar{Q}}_1(t), & t \le T_1 - s \\ 0, & t > T_1 - s \end{cases} \tag{5.20}$$

$$\hat{\bar{Q}}_{2,s}(t) = \begin{cases} \hat{\bar{Q}}_2(t), & t \le T_2 - s \\ 0, & t > T_2 - s \end{cases} \tag{5,21}$$

Let x_i denote the entrance time to the study of the patient with the uncensored observation at t_i. Finally $\delta_i = 1$ or 0 depending on whether the patient with the uncensored observation at t_i was from group 1 ($\delta=1$) or 2 ($\delta=0$). Estimate ℓ_{1i}, ℓ_{2i} by:

$$\hat{\ell}_{1i} = R_{1i} + \sum_{j=1}^{i-1} \delta_j \frac{\hat{\bar{Q}}_{1,x_j}(t_i)}{\hat{\bar{Q}}_{1,x_j}(t_j)} \tag{5.22}$$

$$\hat{\ell}_{2i} = R_{2i} + \sum_{j=1}^{i-1} (1-\delta_j) \frac{\hat{\bar{Q}}_{2,x_j}(t_i)}{\hat{\bar{Q}}_{2,x_j}(t_j)} \tag{5.23}$$

The values $\hat{\ell}_{1i}$, $\hat{\ell}_{2i}$ are then substituted into the expression for $\hat{c}^2$ (5.5).

In the above Q_1 and Q_2 follow the earlier mentioned convention of left rather than right continuous cdfs.

6. Comments and Additions.

(6.1) Consider the class of test statistics:

$$T_g = \sum_{\ell=1}^{K} g(R_i) \left(\delta_i - \frac{R_{1i}}{R_i} \right) \tag{6.2}$$

The function g is chosen to achieve good local power against specific types of alternatives. The choice $g(R) = R$ corresponds to Gehan's [6] generalization of the Wilcoxon test. Our remarks concerning σ^2_{MI} and σ^2_p, found in sections 3 and 4, carry over to T_g. The estimators $\hat{\sigma}^2$ and σ^{*2} are now defined by:

$$\hat{\sigma}_g^2 = \sum_{i=1}^{K} \frac{\ell_{1i}\ell_{2i}(R_i-1)}{\ell_i(\ell_i-1)R_i} \, g^2(R_i) \tag{6.3}$$

$$\sigma_g^{*2} = \sum_{i=1}^{K} \frac{\hat{\ell}_{1i}\hat{\ell}_{2i}(R_i-1)}{\hat{\ell}_i(\hat{\ell}_i-1)R_i} \, g^2(R_i) \tag{6.4}$$

For the case $g(r) = \sqrt{r}$ use of (4.10) gives the following expression for the variance of T_g under known censoring times:

$$\int \ell_1(u)\ell_2(u)\bar{F}(u)dF(u) \tag{6.5}$$

Taking expectations in (6.4) gives the variance under random censoring:

$$nm\int \bar{H}(u)\bar{I}(u)\bar{F}(u)dF(u) \tag{6.6}$$

For the case $g(r) = r$ the variance under known censoring times is given by:

$$\int \ell_1(u)\ell_2(u)\bar{F}(u)[(\ell(u)-2)\bar{F}(u)+2]dF(u) \tag{6.7}$$

The variance under random censoring equals:

(6.8) $$nm\int \bar{H}\,\bar{I}\,\bar{F}_{\cdot}\,[(n-1)\bar{F}\,\bar{H}+(m-1)\bar{F}\,\bar{I}+2]\,dF$$

Working from (6.5) and (6.7) by estimating these expressions yield estimators which are practically the same as σ_g^{*2}.

(6.9) The MH approach as extended by Mantel [8] combines a class of dependent chi-square tables by summing the variances of appropriately defined uncorrelated random variables across the several tables. I believe that the same behavior of the variance estimator discussed in section 3 occurs more generally. For a simple example suppose that $N_0 = n$ and the conditional distribution of N_i given N_{i-1} is $B(N_{i-1},p)$, $i = 1, 2, \ldots$.

We wish to test $p = p_0$ vs $p < p_0$. The appropriate version of the MH statistic would be:

(6.10) $$T = \frac{\sum_i (N_i - N_{i-1}P_0)}{\sqrt{p_0 q_0 \sum_i N_{i-1}}} = \frac{Mq_0 - n}{\sqrt{Mp_0 q_0}}$$

where $M = \sum_i N_i$. H_0 will be rejected for T negative and large in absolute value equivalently for -T large and positive. But the numerator of -T, $n-Mq_0$, is large when M is small, in which case the denominator $\sqrt{Mp_0q_0}$ is small. In this case $n-Mq_0$ and Mp_0q_0 have correlation -1.

The true distribution of M is negative binomial with parameters n and q_0:

(6.11) $$P_r(M=r) = \binom{r-1}{n-1} p_0^{n-r} q_0^r, \quad r \geq n$$

The variance of $Mq_0 - n$ under H_0 equals np_0; its unbiased MH variance estimator is Mp_0q_0.

The distortion of P values in this case is serious even for fairly large values of n.

The MH procedure is clever and useful, but should be used with care.

References

(1) Cox, D.R. (1972). "Regression models and life tables." J.R. Statist. Soc. B, 34, 187-220.

(2) Birnbaum, A. (1962). "On the foundations of statistical inference." JASA, 57, 296-326.

(3) Cox, D.R. (1975). "Partial liklihood." Biometrika, 62, 269-76.

(4) Efron, B. (1967). "The two sample problem with censored data." Proc. 5th Berkeley Symp., 4, 831-53.

(5) Efron, B. (1977). "The efficiency of Cox's liklihood function for censored data." J. Amer. Statist. Assoc., 72, 557-65.

(6) Gehan, E.A. (1965). "A generalized Wilcoxon test for comparing arbitrarily single-censored samples." Biometrika, 52, 203-24.

(7) Mantel, N. and Haenszel, W. (1959). "Statistical aspects of the analysis of data from retrospective studies of disease." J. Nat. Cancer Inst., 22, 719-28.

(8) Mantel, N. (1963). "Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure." J. Am. Statist. Assoc., 58, 690-700.

(9) Mantel, N. (1966). "Evaluation of survival data and two new rank order statistics arising in its consideration." Cancer Chemotherapy Reports, 50, 163-170.

(10) Peto, R. and Peto, J. (1972). "Asymptotically efficient rank invariant test procedures." J.R. Statist. Soc. A, 135, 185-206.

(11) Peto, R. (1972). "Rank tests of maximal power against Lehman-type alternatives." Biometrika, 59, 472-75.